HW1

Hao-Chun Chen

2024-09-17

Table of contents

```
library(Hmisc)
```

```
Attaching package: 'Hmisc'

The following objects are masked from 'package:base':

    format.pval, units
```

```
library(palmerpenguins)
latex(describe(penguins_raw), file = "", caption.placement = "top")
```

penguins_raw
17 Variables 344 Observations

studyName

n	missing	distinct
344	0	3

Value	PAL0708	PAL0809	PAL0910
Frequency	110	114	120
Proportion	0.320	0.331	0.349

Sample Number

n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
344	0	152	1	63.15	46.35	6.15	12.00	29.00	58.00	95.25	121.00	134.85

lowest : 1 2 3 4 5, highest: 148 149 150 151 152

Species

n	missing	distinct
344	0	3

Value	Adelie Penguin (Pygoscelis adeliae)	Chinstrap penguin (Pygoscelis antarctica)	Gentoo penguin (Pygoscelis papua)
Frequency	152	68	124
Proportion	0.442	0.198	0.360

Region

n	missing	distinct	value
344	0	1	Anvers

Value	Anvers
Frequency	344
Proportion	1

Island

n	missing	distinct
344	0	3

Value	Biscoe	Dream	Torgersen
Frequency	168	124	52
Proportion	0.488	0.360	0.151

Stage

n	missing	distinct	value
344	0	1	Adult, 1 Egg Stage

Value	Adult, 1 Egg Stage
Frequency	344
Proportion	1

Individual ID

n	missing	distinct
344	0	190

lowest : N100A1 N100A2 N10A1 N10A2 N11A1 , highest: N98A2 N99A1 N99A2 N9A1 N9A2

Clutch Completion

n	missing	distinct
344	0	2

Value	No	Yes
Frequency	36	308
Proportion	0.105	0.895

Date Egg

```
        n    missing   distinct       Info       Mean        Gmd        .05        .10
      344          0         50      0.999 2008-11-27        328 2007-11-12 2007-11-16
      .25        .50        .75        .90        .95
2007-11-28 2008-11-09 2009-11-16 2009-11-22 2009-11-26

lowest : 2007-11-09 2007-11-10 2007-11-11 2007-11-12 2007-11-13
highest: 2009-11-22 2009-11-23 2009-11-25 2009-11-27 2009-12-01
```

Culmen Length (mm)

n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
342	2	164	1	43.92	6.274	35.70	36.60	39.23	44.45	48.50	50.80	51.99

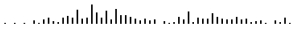
```
lowest : 32.1 33.1 33.5 34   34.1, highest: 55.1 55.8 55.9 58   59.6
```

Culmen Depth (mm)

n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
342	2	80	1	17.15	2.267	13.9	14.3	15.6	17.3	18.7	19.5	20.0

```
lowest : 13.1 13.2 13.3 13.4 13.5, highest: 20.7 20.8 21.1 21.2 21.5
```

Flipper Length (mm)



n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
342	2	55	0.999	200.9	16.03	181.0	185.0	190.0	197.0	213.0	220.9	225.0

```
lowest : 172 174 176 178 179, highest: 226 228 229 230 231
```

Body Mass (g)

n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
342	2	94	1	4202	911.8	3150	3300	3550	4050	4750	5400	5650

```
lowest : 2700 2850 2900 2925 2975, highest: 5850 5950 6000 6050 6300
```

Sex

n	missing	distinct
333	11	2

Value	FEMALE	MALE
Frequency	165	168
Proportion	0.495	0.505

Δ 15 N (o/oo):

n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
330	14	330	1	8.733	0.6323	7.897	8.047	8.300	8.652	9.172	9.491	9.689

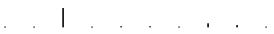
```
lowest : 7.6322  7.63452 7.63884 7.68528 7.6887 , highest: 9.93727 9.98044 10.0202 10.0237 10.0254
```

Δ 13 C (o/oo):

n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
331	13	331	1	-25.69	0.9093	-26.79	-26.69	-26.32	-25.83	-25.06	-24.53	-24.36

```
lowest : -27.0185 -26.9547 -26.8964 -26.8648 -26.8635, highest: -24.1657 -24.1026 -23.9031 -23.8902 -23.7877
```

Comments



n	missing	distinct
54	290	10

```
lowest : Adult not sampled.                                                    Adult not sampled. Nest never observed with ful
highest: No blood sample obtained.                                             No delta15N data received from lab.
```

```
library(Hmisc)
library(DataExplorer)
latex(describe(penguins_raw), file = "", caption.placement = "top")
```

penguins_raw
17 Variables 344 Observations

---

studyName

n	missing	distinct
344	0	3

```
Value      PAL0708 PAL0809 PAL0910
Frequency      110     114     120
Proportion   0.320   0.331   0.349
```

Sample Number

n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
344	0	152	1	63.15	46.35	6.15	12.00	29.00	58.00	95.25	121.00	134.85

```
lowest :   1   2   3   4   5, highest: 148 149 150 151 152
```

Species

n	missing	distinct
344	0	3

```
Value           Adelie Penguin (Pygoscelis adeliae) Chinstrap penguin (Pygoscelis antarctica)
Frequency                                       152                                        68
Proportion                                    0.442                                     0.198

Value             Gentoo penguin (Pygoscelis papua)
Frequency                                       124
Proportion                                    0.360
```

Region

n	missing	distinct	value
344	0	1	Anvers

Value	Anvers
Frequency	344
Proportion	1

Island

n	missing	distinct
344	0	3

Value	Biscoe	Dream	Torgersen
Frequency	168	124	52
Proportion	0.488	0.360	0.151

Stage

n	missing	distinct	value
344	0	1	Adult, 1 Egg Stage

Value	Adult, 1 Egg Stage
Frequency	344
Proportion	1

Individual ID

n	missing	distinct
344	0	190

lowest : N100A1 N100A2 N10A1 N10A2 N11A1 , highest: N98A2 N99A1 N99A2 N9A1 N9A2

Clutch Completion

n	missing	distinct
344	0	2

Value	No	Yes
Frequency	36	308
Proportion	0.105	0.895

Date Egg

n	missing	distinct	Info	Mean	Gmd	.05	.10
344	0	50	0.999	2008-11-27	328	2007-11-12	2007-11-16
.25	.50	.75	.90	.95			
2007-11-28	2008-11-09	2009-11-16	2009-11-22	2009-11-26			

lowest : 2007-11-09 2007-11-10 2007-11-11 2007-11-12 2007-11-13
highest: 2009-11-22 2009-11-23 2009-11-25 2009-11-27 2009-12-01

Culmen Length (mm)

n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
342	2	164	1	43.92	6.274	35.70	36.60	39.23	44.45	48.50	50.80	51.99

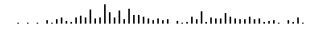
```
lowest : 32.1 33.1 33.5 34   34.1, highest: 55.1 55.8 55.9 58   59.6
```

Culmen Depth (mm)

n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
342	2	80	1	17.15	2.267	13.9	14.3	15.6	17.3	18.7	19.5	20.0

```
lowest : 13.1 13.2 13.3 13.4 13.5, highest: 20.7 20.8 21.1 21.2 21.5
```

Flipper Length (mm)



n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
342	2	55	0.999	200.9	16.03	181.0	185.0	190.0	197.0	213.0	220.9	225.0

```
lowest : 172 174 176 178 179, highest: 226 228 229 230 231
```

Body Mass (g)

n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
342	2	94	1	4202	911.8	3150	3300	3550	4050	4750	5400	5650

```
lowest : 2700 2850 2900 2925 2975, highest: 5850 5950 6000 6050 6300
```

Sex

n	missing	distinct
333	11	2

Value	FEMALE	MALE
Frequency	165	168
Proportion	0.495	0.505

Δ 15 N (o/oo):

n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
330	14	330	1	8.733	0.6323	7.897	8.047	8.300	8.652	9.172	9.491	9.689

```
lowest : 7.6322  7.63452 7.63884 7.68528 7.6887 , highest: 9.93727 9.98044 10.0202 10.0237 10.0254
```

Δ 13 C (o/oo):

n	missing	distinct	Info	Mean	Gmd	.05	.10	.25	.50	.75	.90	.95
331	13	331	1	-25.69	0.9093	-26.79	-26.69	-26.32	-25.83	-25.06	-24.53	-24.36

```
lowest : -27.0185 -26.9547 -26.8964 -26.8648 -26.8635, highest: -24.1657 -24.1026 -23.9031 -23.8902 -23.7877
```

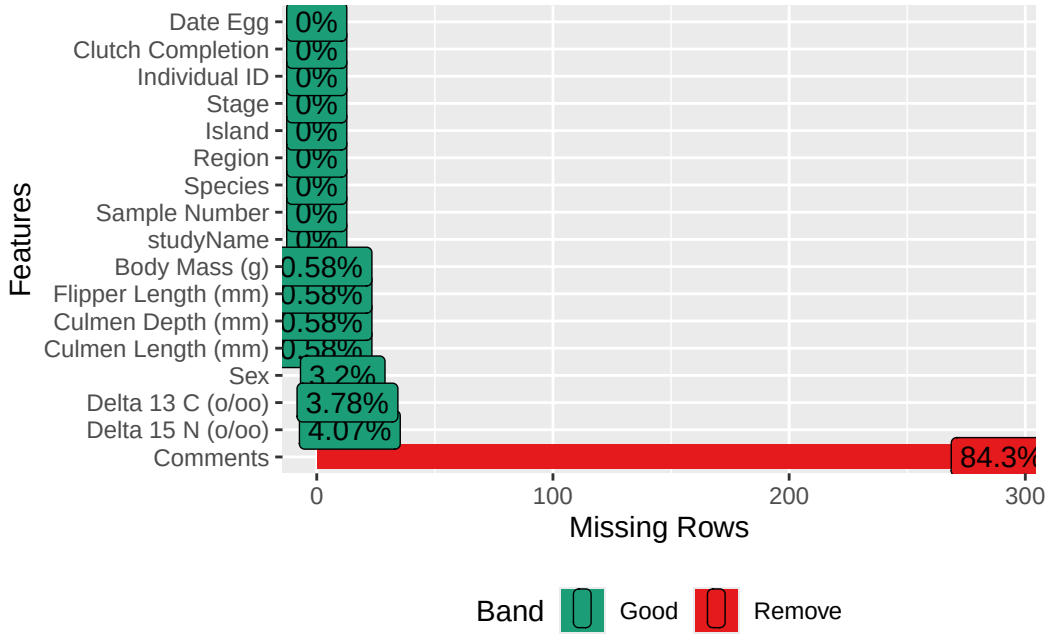
Comments

n	missing	distinct
54	290	10

```
lowest : Adult not sampled.
highest: No blood sample obtained.
```

```
Adult not sampled. Nest never observed with ful
No delta15N data received from lab.
```

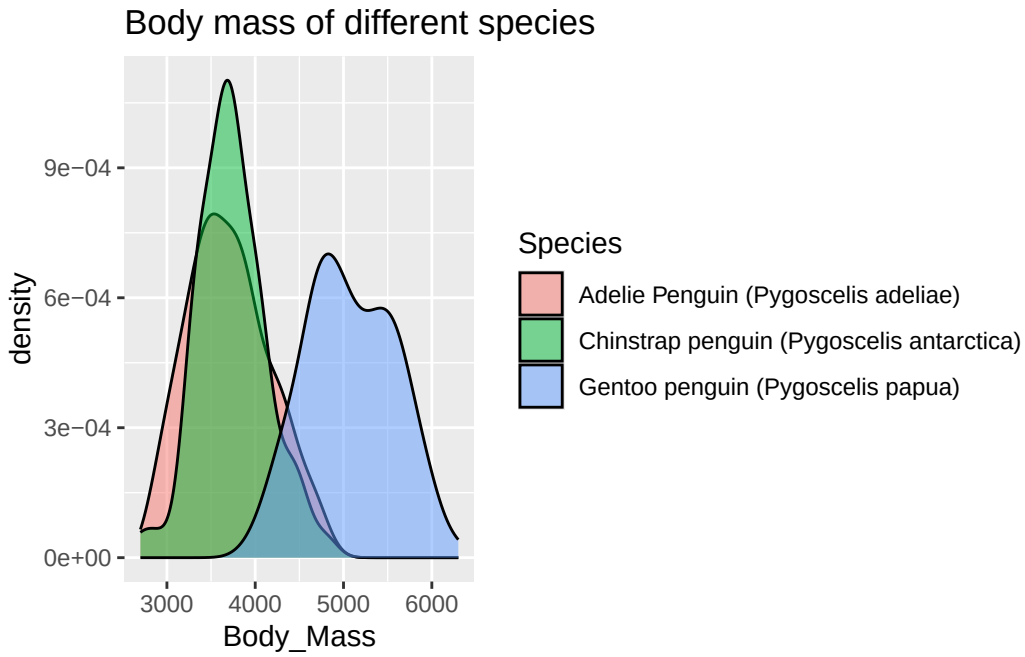
```
plot_missing(penguins_raw)
```



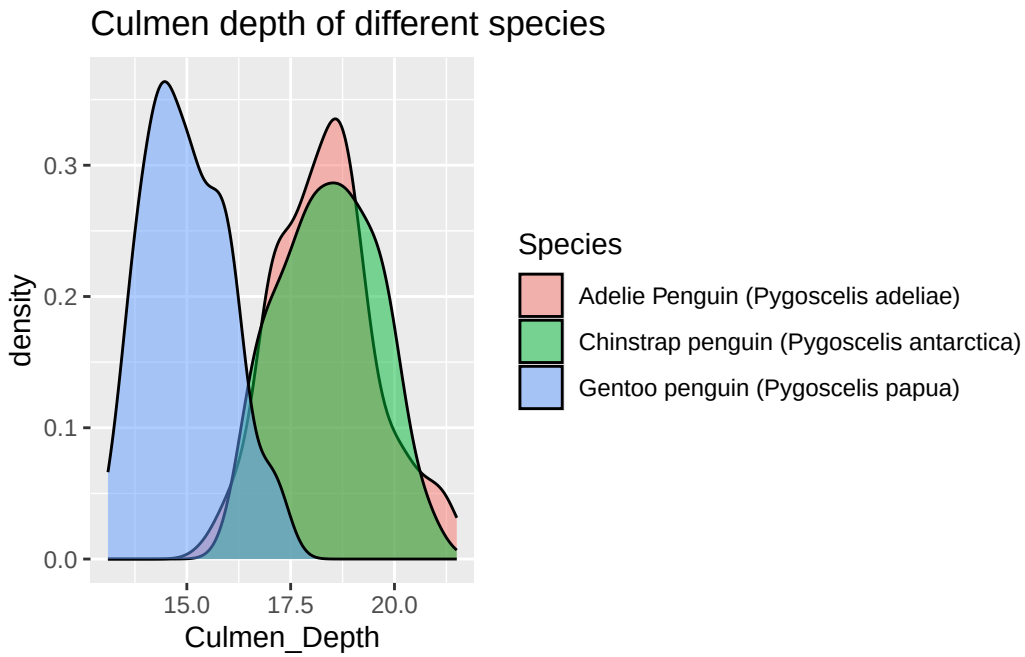

```
library(dplyr)
library(ggplot2)
penguins_raw <- penguins_raw %>%
  rename(
    Culmen_Depth = `Culmen Depth (mm)`,
    Culmen_Length = `Culmen Length (mm)`,
    Flipper_Length = `Flipper Length (mm)`,
    Body_Mass = `Body Mass (g)`
  )

penguin_subset <- penguins_raw %>%
  select(Species, Culmen_Depth, Culmen_Length, Flipper_Length, Body_Mass) %>%
  filter(!is.na(Species),
```

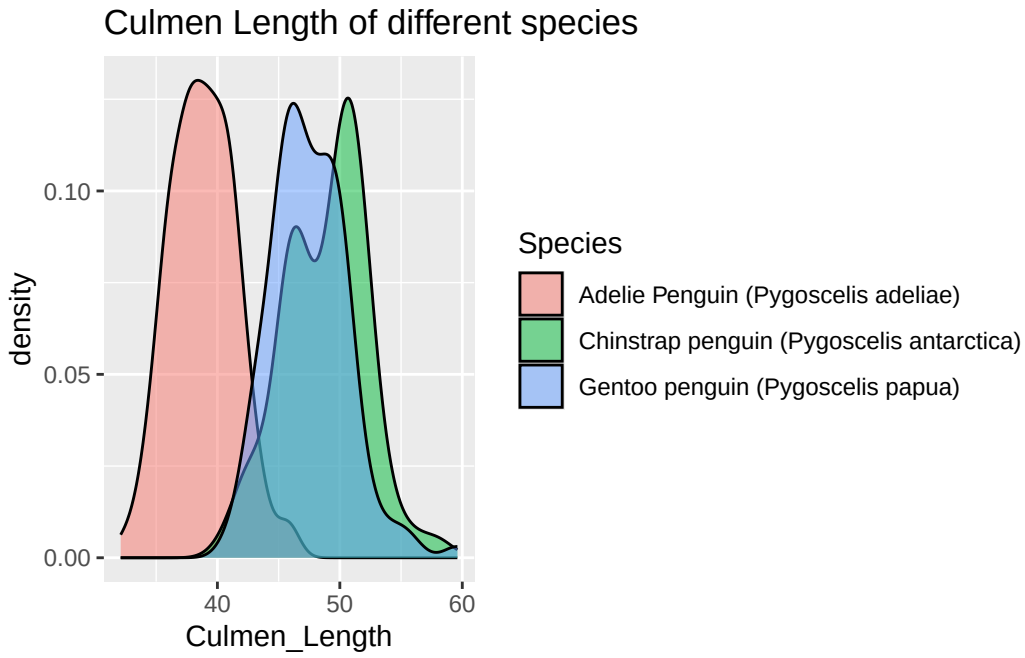
```
         !is.na(Culmen_Depth),
         !is.na(Culmen_Length),
         !is.na(Flipper_Length),
         !is.na(Body_Mass))
par(mfrow=c(2,2))
ggplot(penguin_subset)+
  geom_density(mapping = aes(x=Body_Mass,fill=Species),alpha=0.5)+
  ggtitle("Body mass of different species")
```



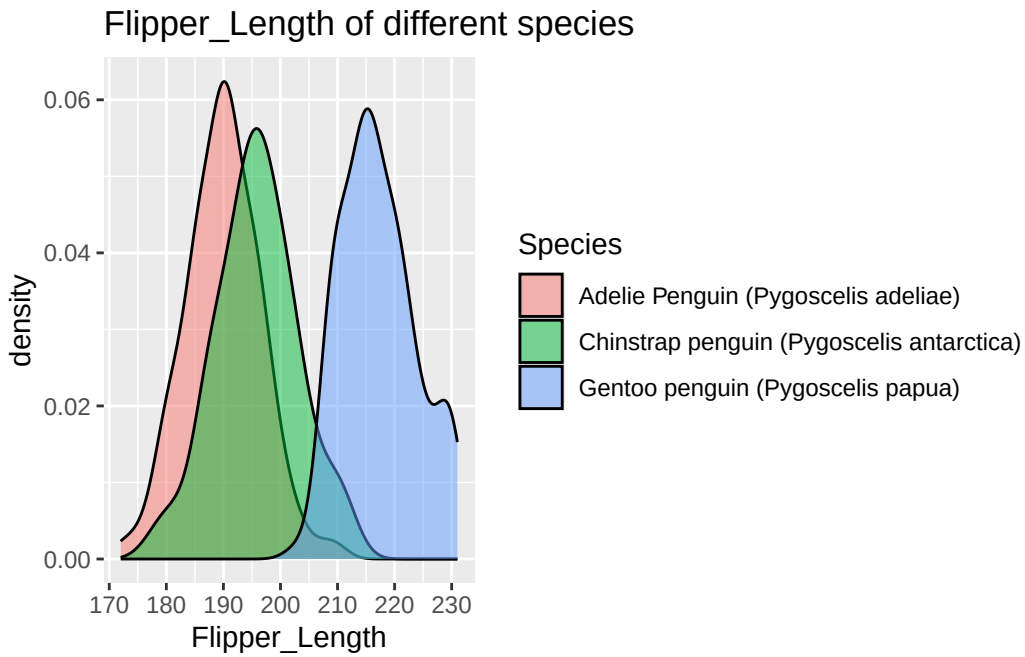

```
ggplot(penguin_subset)+
  geom_density(mapping = aes(x=Culmen_Depth,fill=Species),alpha=0.5)+
  ggtitle("Culmen depth of different species")
```




```
ggplot(penguin_subset)+
  geom_density(mapping = aes(x=Culmen_Length,fill=Species),alpha=0.5)+
  ggtitle("Culmen Length of different species")
```



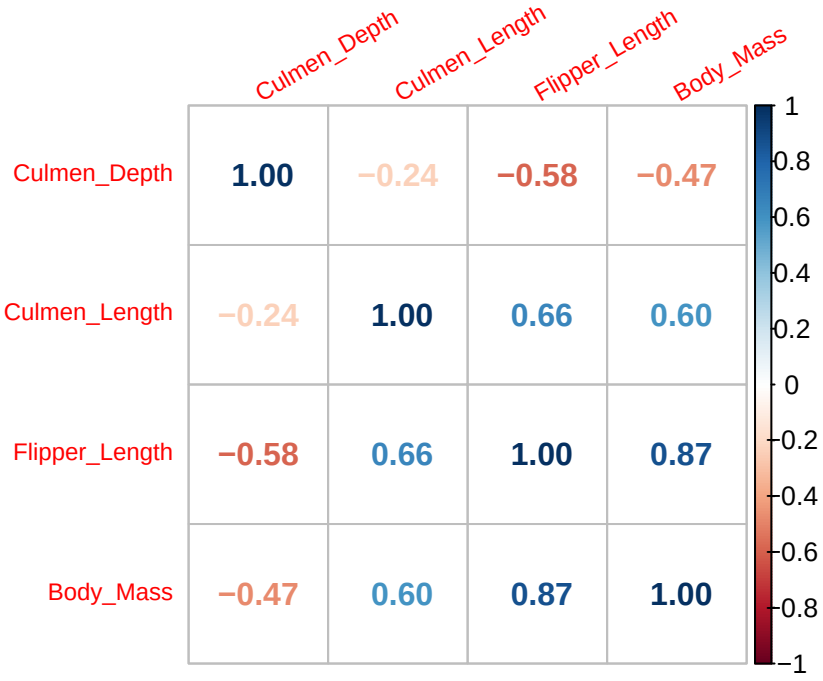
```
ggplot(penguin_subset)+
  geom_density(mapping = aes(x=Flipper_Length,fill=Species),alpha=0.5)+
  ggtitle("Flipper_Length of different species")
```




```
library(corrplot)
```

```
corrplot 0.92 loaded
```

```
corrplot(
  cor(penguin_subset[,-1]),
  method = "number",
  type = "full",
  tl.col = "red",
  tl.cex = 0.75,
  tl.srt = 30
)
```


Culmen_Depth
Culmen_Length
Flipper_Length
Body_Mass
Culmen_Depth
Culmen_Length
Flipper_Length
Body_Mass
1.00
−0.24
−0.58
−0.47
−0.24
1.00
0.66
0.60
−0.58
0.66
1.00
0.87
−0.47
0.60
0.87
1.00
1
0.8
0.6
0.4
0.2
0
−0.2
−0.4
−0.6
−0.8
−1